

IN THE UNITED STATES DISTRICT COURT  
FOR THE SOUTHERN DISTRICT OF TEXAS  
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*,

Plaintiffs,

v.

RICK PERRY, *et al.*,

Defendants.

Civil Action No. 2:13-cv-193 (NGR)

UNITED STATES OF AMERICA,

Plaintiff,

TEXAS LEAGUE OF YOUNG VOTERS  
EDUCATION FUND, *et al.*,

Plaintiff-Intervenors,

TEXAS ASSOCIATION OF HISPANIC  
COUNTY JUDGES AND COUNTY  
COMMISSIONERS, *et al.*,

Plaintiff-Intervenors,

v.

STATE OF TEXAS, *et al.*,

Defendants.

Civil Action No. 2:13-cv-263 (NGR)



TEXAS STATE CONFERENCE OF NAACP  
BRANCHES, *et al.*,

Plaintiffs,

v.

NANDITA BERRY, *et al.*,

Defendants.

Civil Action No. 2:13-cv-291 (NGR)

BELINDA ORTIZ, *et al.*,

Plaintiffs,

v.

STATE OF TEXAS, *et al.*,

Defendants.

**DEFENDANTS' OBJECTIONS AND RESPONSES TO PLAINTIFFS  
AND PLAINTIFF-INTERVENORS' FIRST SET OF INTERROGATORIES**

**TO:** All Plaintiffs and Plaintiff-Intervenors, by and through their attorneys of record.

Pursuant to Rule 33 of the Federal Rules of Civil Procedure, the State of Texas, Rick Perry, John Steen and Steve McCraw, by and through the Attorney General for the State of Texas, serve these Objections and Responses to Plaintiffs and Plaintiff-Intervenors' First Set of Interrogatories.

**GENERAL OBJECTIONS**

Defendants object to each interrogatory: (1) insofar as it seeks information not in Defendants' possession, custody, or control; (2) insofar as it seeks information that was prepared for or in anticipation of litigation, constitutes attorney work product, contains attorney-client communications, or is otherwise privileged; (3) insofar as it seeks information which is publicly available or otherwise equally available and/or uniquely or equally available from third parties; (4) insofar as it seeks information that does not specifically refer to the events which are the subject matter of this litigation; and (5) insofar as it seeks information not relevant to the subject matter of this litigation nor reasonably calculated to lead to the discovery of admissible evidence.

These responses and objections are made on the basis of information now known to Defendants and are made without waiving any further objections to, or admitting the relevancy or materiality of, any of the information requested. Defendants' investigation, discovery, and preparation for proceedings are continuing and all answers are given without prejudice to Defendants' right to introduce or object to the discovery of any documents, facts, or information

from 2010 through June 2013 were the responsibility of Rebecca Davio, then the Assistant Director of the Driver License Division and her executive management team.

**28. Identify all Texas State and local public agencies and entities that issue or are authorized to issue photographic identification and state the eligibility requirements for obtaining a form of photographic identification from each such agency or entity.**

**OBJECTION:** Defendants object to this interrogatory on the ground that it contains two separate interrogatories. By asking Defendants to “identify all Texas State and local public agencies and entities that issue or are authorized to issue photographic identification” and then asking Defendants to “state the eligibility requirements for obtaining a form of photographic identity from each such agency or entity”, the interrogatory “introduces a line of inquiry that is separate and distinct from the inquiry made by the portion of the interrogatory that precedes it”. *Willingham v. Ashcroft*, 226 F.R.D. 57, 59 (D.D.C. 2005) (internal quotations and citations omitted). Defendants will therefore construe Interrogatory No. 28 as two separate interrogatories.

Defendants object to this interrogatory as overly broad, unduly burdensome, and vague insofar as it fails to define the term “photo identification” and seeks to impose an obligation on Defendants to identify every document that might fit within the definition of that term without regard to the information necessary to obtain such document or the information contained in such document. Further, the interrogatory seeks information that is not relevant and is not likely to lead to the discovery of relevant or admissible evidence. Defendants further object to this interrogatory to the extent it seeks information not personally known or controlled by the



Defendants but is known and controlled by third parties, including independent officers of the state and local governments. Defendants further object to this interrogatory to the extent it seeks information that is publicly available or equally accessible to the plaintiffs and plaintiff-intervenors.

Subject to and without waiving the foregoing objections, Defendants respond generally as follows:

**Interrogatory 28a**

The Texas Department of Public Safety issues the following forms of identification containing photographs:

- Texas Driver License
- Texas Personal Identification Certificate
- Texas Concealed Handgun License
- Election Identification Certificate
- Department of Public Safety Employee Identification Cards
- DPS and Capitol Complex employee access cards
- Private Security Pocket Card

The agencies listed below use DPS resources to issue security access cards to their employees:

Texas Senate

Texas House of Representatives

Texas Legislative Council

Legislative Budget Board

Legislative Reference Library

Sunset Advisory Commission  
Supreme Court of Texas  
Texas Board of Law Examiners  
Court of Criminal Appeals  
Office of Court Administration  
State Prosecuting Attorney  
Office of Capital Writs  
Third District Court of Appeals  
State Commission on Judicial Conduct  
Texas State Law Library  
Office of the Governor  
Office of the Attorney General  
Texas Facilities Commission  
Comptroller of Public Accounts  
General Land Office and Veteran Land Board  
Texas State Library and Archives  
Texas Secretary of State  
State Auditor's Office  
State Securities Board  
Department of Information Resources  
Texas Workforce Commission  
Fire Fighters' Pension Commission

Texas Employees' Retirement System  
Texas Real Estate Commission  
Texas Department of Housing and Community Affairs  
Texas Board of Tax Professionals  
Pension Review Board  
Public Finance Authority  
Bond Review Board  
Texas Aerospace Commission  
Texas Ethics Commission  
Public Insurance Council  
State Office of Administrative Hearings  
Health Professions Council  
Veterans Commission  
Texas Department of Public Safety  
Commission on Jail Standards  
Texas Commission on Fire Protection  
Texas Savings and Loan Department  
Department of Banking  
Department of Licensing and Regulation  
Texas Department of Insurance  
Texas Railroad Commission  
Texas State Board of Public Accountancy

Texas Board of Architectural Examiners  
Office of Consumer Credit Commission  
Structural Pest Control Board  
Public Utility Commission  
Public Utility Council  
Advisory Commission on State Emergency  
Office of Risk Management  
Board of Professional Geo-Scientist  
Board of Medical Examiners  
Board of Dental Examiners  
Board of Nursing  
Board of Chiropractic Examiners  
Board of Podiatry Examiners  
Funeral Communications  
Optometry Board  
Board of Pharmacy  
Board of Psychologists Examiners  
Health and Human Services Commission  
Executive Council of Physical and Occupational Therapy  
Department of State Health Services  
Department of Assistance and Rehab Services  
Department of Agriculture



Veterinary Medical Examiners  
Water Development Board  
Texas Department of Transportation  
Texas Department of Motor Vehicles  
Texas Juvenile Justice Department  
Department of Criminal Justice  
Department of Criminal Justice-Pardons and Paroles Division  
Texas Education Agency  
Texas Southern University  
LBJ School of Public Affairs  
Office of Demographers  
Texas Forensic Scientist Commission  
Stephen F. Austin State University  
University of North Texas System  
Board of Regents, Texas State University System  
Texas Parks & Wildlife Department  
Texas Historical Commission  
State Preservation Board  
Texas Commission on the Arts  
Treasury Safekeeping Trust Company

These cards contain a digital image of the employee. The cards issued to employees of these agencies consist of a Security Access Card (HID Proximity Card, approved state employees for

access to buildings administered by DPS) and Texas House of Representatives Capitol Access Card (HID Proximity Card, approved House employees for access to House-controlled areas of the state Capitol building under agreement with the State Preservation Board).

**Interrogatory 28b**

With respect to the "eligibility requirements for obtaining" the above listed identification, see the Defendants' responses to Interrogatory Numbers 4a, 4b, 4c, 5a, 5b, 5c, 6a, 6b, 6c, 7a, 7b and 7c. Further, with respect to security access cards issued to employees, the eligibility requirements would be determined by the terms of employment of each agency.

Dated: February 28, 2014

Respectfully submitted,

GREG ABBOTT  
Attorney General of Texas

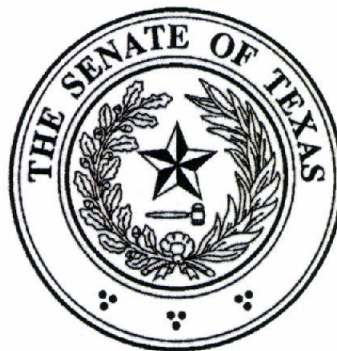
DANIEL T. HODGE  
First Assistant Attorney General

JONATHAN F. MITCHELL  
Solicitor General

J. REED CLAY, JR.  
Special Assistant and Senior Counsel  
to the Attorney General  
Southern District of Texas No. 1160600

/s/ John B. Scott

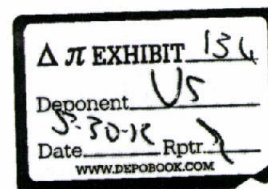
JOHN B. SCOTT  
Deputy Attorney General for Litigation  
Southern District of Texas No. 10418  
ATTORNEY-IN-CHARGE



# SENATE RULES

adopted by  
82nd LEGISLATURE  
January 19, 2011

Senate Resolution No. 36



Rule 5.10

---

A House Concurrent Resolution taken up in its calendar order on a House bill day may not be further considered when a Senate bill day arrives (71 S.J. 1 C.S. 73 (1989)).

When rules have been suspended to permit consideration of a Senate bill on a House bill day, an additional suspension is not required to permit consideration to continue when a Senate bill day arrives (73 S.J. Reg. 1082 (1993)).

**SPECIAL ORDERS**

**Rule 5.11.** (a) Any bill, resolution, or other measure may on any day be made a special order for a future time of the session by an affirmative vote of two-thirds of the members present.

(b) A special order shall be considered at the time for which it is set and considered from day to day until disposed of, unless at the time so fixed there is pending business under a special order, but such pending business may be suspended by a two-thirds vote of all the members present. If a special order is not reached or considered at the time fixed, it shall not lose its place as a special order. All special orders shall be subject to any Joint Rules and Rule 5.10.

(c) Upon the affirmative vote of four-fifths of the members present, a special order may be reset to an earlier time than previously scheduled.

(d) Notwithstanding Subsection (a) of this rule, a bill or resolution relating to voter identification requirements reported favorably from the Committee of the Whole Senate may be set as a special order for a time at least 24 hours after the motion is adopted by a majority of the members of the Senate.

**Editorial Notes**

A bill once set as a special order does not lose its place on the calendar of special orders if not taken up at the hour for which it is set.

A special order, the hour for the consideration of which has arrived, takes precedence of the unfinished business unless the unfinished business is itself a special order.

IN THE UNITED STATES DISTRICT COURT  
FOR THE SOUTHERN DISTRICT OF TEXAS  
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*,

Plaintiffs,

v.

RICK PERRY, *et al.*,

Defendants.

Civil Action No. 2:13-cv-193 (NGR)

UNITED STATES OF AMERICA,

Plaintiff,

TEXAS LEAGUE OF YOUNG VOTERS  
EDUCATION FUND, *et al.*,

Plaintiff-Intervenors,

TEXAS ASSOCIATION OF HISPANIC  
COUNTY JUDGES AND COUNTY  
COMMISSIONERS, *et al.*,

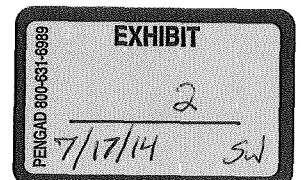
Plaintiff-Intervenors,

v.

STATE OF TEXAS, *et al.*,

Defendants.

Civil Action No. 2:13-cv-263 (NGR)





TEXAS STATE CONFERENCE OF NAACP  
BRANCHES, *et al.*,

Plaintiffs,

v.

JOHN STEEN, *et al.*,

Defendants.

Civil Action No. 2:13-cv-291 (NGR)

BELINDA ORTIZ, *et al.*,

Plaintiffs,

v.

STATE OF TEXAS, *et al.*,

Defendants

Civil Action No. 2:13-cv-348 (NGR)

#### DECLARATION OF MICHAEL MIMS

I, Michael Mims, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as Program Manager for HealthCare Identity Management for the Veterans Health Administration of the U.S. Department of Veterans Affairs. I have served in this capacity for April 2012. My duties include the maintenance and validation of the identity data stored within the Master Veteran Index.
2. I have personal knowledge of the information contained in this declaration based upon my work for the Veterans Health Administration of the United States Department of Veterans Affairs ("VHA"), and my personal involvement in supervision of the database comparison described below.

3. On March 11, 2014, my agency received an encrypted hard drive containing data regarding Texas registered voters from the United States Department of Justice in the form of a 7 gigabyte .csv file containing 13,564,420 records for Texas registered voters. This file was corrupted in the process of loading the data and so an identical dataset was delivered to VHA on March 26, 2014. VHA was able to load all 13,564,420 records.

4. The Texas data that VHA received contained the following fields:

- Vuid	- AGDNlast_TAG
- last_name	- AGDNlast_unique
- first_name	- AGD
- middle_name	- AGD_TAG
- ssn	- AGD_unique
- ssn4_str	- ADN
- gender_code	- ADN_TAG
- official_id	- ADN_unique
- dl_id_str	- AGN
- date_of_birth	- AGN_TAG
- voter_status_code	- AGN_unique
- block_number	- GDN
- street_name	- GDN_TAG
- city	- GDN_unique
- state	- DLN
- zip_code	- DLN_TAG
- zip_str	- DLN_unique
- mail_adrs_1	- unique_count
- last_fix	- FMLD
- multi_lastname1	- FMLD_TAG
- multi_lastname2	- FMLD_unique
- first_fix	- SDZ
- middle_initial	- SDZ_TAG
- fem_str	- SDZ_unique
- dob_str	- SND
- zip5_res_alt	- SND_TAG
- str_num_res_alt	- SND_unique
- AGDN	- FML1D
- AGDN_TAG	- FML1D_TAG
- AGDN_unique	- FML1D_unique
- AGDNlast	- FML2D

- |                 |      |
|-----------------|------|
| - FML2D_TAG     | - K2 |
| - FML2D_unique  | - K3 |
| - unique_count2 | - L2 |
| - var_fill      | - L3 |

5. From the fields listed above, VHA loaded the following variables that were specifically used in completing the matching process:

- |            |                  |
|------------|------------------|
| - void     | - last_fix       |
| - GDN      | - first_fix      |
| - FMLD     | - middle_initial |
| - FML1D    | - dob_str        |
| - FML2D    | - SDZ            |
| - AGDN     | - SND            |
| - AGDNlast | - var_fill       |
| - AGD      | - K2             |
| - ADN      | - K3             |
| - AGN      | - L2             |
| - SSN      | - L3             |
| - ssn4_str |                  |

6. VHA executed the data preparation and comparison steps, attached as Ex. A, that were provided by the United States Department of Justice.
7. Through this process, VHA attempted to match particular combinations of identifying information for Texas registered voters (*e.g.*, first and last name, gender, and date of birth) with the same or related combinations of identifying information with respect to holders of a Veteran Identification Card. For each particular combination, VHA identified instances where the identifying information for a Texas registered voter matched with the same or related combinations of identifying information for one or more holders of a Veteran Identification Card. The frequencies of any missing values in the underlying VHA dataset are provided in Ex. B.
8. For each sweep through the relevant VHA dataset, we appended a column to the Texas data that VHA loaded to indicate, on a record-by-record basis, the output of the database

comparisons on all of the requested combinations where there was one or more matches.

In total, out of the 13,564,420 records available to be matched, VHA found matches on one or more combinations for 299,320 voter records.

9. On April 30, 2014, VHA transferred responsive data to the Department of Justice, in the form of a 117 megabyte text (.txt) file, with vertical bar delimiters. This file contained the match results for each of the 299,320 records where VHA found one or more matches. The voter file records where no matches were found were not returned.
10. The results of each sweep, as described in Ex. A, can be found in the following columns of the data VHA returned to the Department of Justice:
  - a. "VHA\_\_USA" contains results of the Stage 1, Step 3.1.1, Combination A match;
  - b. "VHA\_\_USB" contains results of the Stage 1, Step 3.1.2, Combination B match;
  - c. "VHA\_\_USC" contains results of the Stage 1, Step 3.1.2, Combination C match;
  - d. "VHA\_\_USD" contains results of the Stage 1, Step 3.1.2, Combination D match;
  - e. "VHA\_\_USE" contains results of the Stage 1, Step 3.1.2, Combination E match;
  - f. "VHA\_\_USF" contains results of the Stage 1, Step 3.1.2, Combination F match;
  - g. "VHA\_\_USG" contains results of the Stage 2, Step 3.2.1, Combination G match;
  - h. "VHA\_\_USH" contains results of the Stage 2, Step 3.2.2, Combination H match;
  - i. "VHA\_\_USI" contains results of the Stage 2, Step 3.2.2, Combination I match;
  - j. "VHA\_\_USSSN" contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;
  - k. "VHA\_\_USK2" contains results of the Stage 2, Step 3.2.3, Combination K to G match;
  - l. "VHA\_\_USK" contains results of the Stage 2, Step 3.2.3, Combination K match;
  - m. "VHA\_\_USK3" contains results of the Stage 2, Step 3.2.3, Combination K to L match;

- n. "VHA\_USL2" contains results of the Stage 2, Step 3.2.4, Combination L to G match;
- o. "VHA\_USL" contains results of the Stage 2, Step 3.2.4, Combination L match;
- p. "VHA\_USL3" contains results of the Stage 2, Step 3.2.4, Combination L to K match;
- q. "VHA\_USF\_NW" contains results of the Stage 3, Step 3.3.1, Combination F match;
- r. "VHA\_USG\_NW" contains results of the Stage 3, Step 3.3.2, Combination G match;
- s. "VHA\_USI\_NW" contains results of the Stage 3, Step 3.3.2, Combination I match;
- t. "VHA\_USSSN\_NW" contains results of the Stage 3, Step 3.3.2, nine-digit social security number match;
- u. "VHA\_USK2\_NW" contains results of the Stage 3, Step 3.3.3, Combination K to G match;
- v. "VHA\_USK\_NW" contains results of the Stage 2, Step 3.3.3, Combination K match;
- w. "VHA\_USK3\_NW" contains results of the Stage 2, Step 3.3.3, Combination K to L match;
- x. "VHA\_USL2\_NW" contains results of the Stage 3, Step 3.3.4, Combination L to G match;
- y. "VHA\_USL\_NW" contains results of the Stage 3, Step 3.3.4, Combination L match;
- z. "VHA\_USL3\_NW" contains results of the Stage 3, Step 3.3.4, Combination L to K match;
- aa. "texas1" contains results of the Stage 4, Step 3.4.1, Sweep 1 match;
- bb. "texas2" contains results of the Stage 4, Step 3.4.2, Sweep 2 match;
- cc. "texas3" contains results of the Stage 4, Step 3.4.3, Sweep 3 match; and
- dd. "texas4" contains results of the Stage 4, Step 3.4.4, Sweep 4 match.



11. No other matching comparisons between the Texas data and data from VHA were undertaken beyond those set forth in Ex. A.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on May 7, 2014.

A handwritten signature in cursive script, appearing to read "Michael J. Martin", written in dark ink.

Exhibit A

**United States v. Texas: Federal Agency Algorithm Summary**

This document summarizes the database matches that both the United States and all other plaintiffs and the State of Texas have requested from Federal Agencies as part of the *Veasey v. Perry/United States v. Texas* litigation (S.D. Tex).

The matching process proceeds in three parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, the United States' one-to-many matches and the State of Texas's many-to-many matches are conducted between databases.

**PART I: DATABASE PREPARATION**

**Stage 1: Extraction of Available Data from Federal Identification & Disability Databases**

- Step 1.1.1:** Extract complete name into separate first name, middle name, and last name fields.
- Step 1.1.2:** Extract date of birth.
- Step 1.1.3:** Extract gender.
- Step 1.1.4:** Extract residential address and mailing address.
- Step 1.1.5:** Extract social security number.
- Step 1.1.6:** Extract Texas driver license number (only if present in Federal database).

**Stage 2: Separate Valid Identification and Disability Records**

- Step 1.2.1:** Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

Exhibit A

- Step 1.2.2:** Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

**Stage 3: Diagnostics**

- Step 1.3.1:** Report the frequency of missing values for each field.
- Step 1.3.2:** Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.
- Step 1.3.3:** Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

**Stage 4: Standardize Last Name**

- Step 1.4.1:** Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.
- Step 1.4.2:** For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.
- Step 1.4.3:** Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.
- Step 1.4.4:** Code all missing values as blank fields.

Exhibit A

**Stage 5: Standardize First Name and Middle Name**

- Step 1.5.1:** Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.
- Step 1.5.2:** Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.<sup>1</sup>
- Step 1.5.3:** Code all missing values as blank fields.

**Stage 6: Standardize Date of Birth**

- Step 1.6.1:** Convert the date of birth to an eight-digit string of MMDDYYYY.
- Step 1.6.2:** Code all missing values as blank fields.

**Stage 7: Standardize Gender**

- Step 1.7.1:** Code gender as a string of 1 for females and 0 for males.
- Step 1.7.2:** Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.
- Step 1.7.3:** If missing values remain, code all missing values as blank fields.

---

<sup>1</sup> The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Exhibit A

**Stage 8: Standardize Address**

- Step 1.8.1:** Convert the residential ZIP code to a string if it is stored as a numeric field.
- Step 1.8.2:** Where the residential address ZIP code is blank, populate that field with the value in the mailing address ZIP code field, if available.<sup>2</sup>
- Step 1.8.3:** Truncate the residential ZIP code field to the first five digits. *E.g.*, <77777-1234> becomes <77777>.
- Step 1.8.4:** Where the residential address field is blank, populate that field with the value in the mailing address field, if available.
- Step 1.8.5:** Where address field containing street address begins with a street number, isolate the street number. *E.g.*, <123 Main Street> becomes <123>.
- Step 1.8.6:** Where the address field begins with recognized strings indicating a mail box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444> becomes <444>.
- Step 1.8.7:** If missing values remain, code all missing values as blank fields.

**Stage 9: Standardize Social Security Number**

- Step 1.9.1:** Convert the social security number to a string if it is stored as a numeric field.
- Step 1.9.2:** Using full social security number, check for invalid SSNs. In the case of invalid SSNs, code as missing. *E.g.*, <123456789> becomes <>.

---

<sup>2</sup> For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.



Exhibit A

**Step 1.9.3:** Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

**Step 1.9.4:** Code all missing values as blank fields.

## **PART II: DATABASE PREPARATION**

### **Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps**

**Step 2.1.1:** Create Combination A: First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.<sup>3</sup>

**Step 2.1.2:** Create Combination B: Last Name + Gender + DOB + Residential ZIP + Residential Street Number.

**Step 2.1.3:** Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.

**Step 2.1.4:** Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.

**Step 2.1.5:** Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.

**Step 2.1.6:** Create Combination F: First Name + Last Name + Gender + DOB.

**Step 2.1.7:** Create Combination M: Texas Driver License Number (where available).

---

<sup>3</sup> For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Exhibit A

**Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps**

**Step 2.2.1:** Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.<sup>4</sup>

**Step 2.2.2:** Create Combination H: SSN4 + Date of Birth + Residential ZIP.

**Step 2.2.3:** Create Combination I: SSN4 + First Name + Last Name + Date of Birth.

**Step 2.2.4:** Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.

**Step 2.2.5:** Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.

**Step 2.2.6:** Full Social Security Number.

**Stage 3: Construct Identifiers Used Only For Texas's Many-to-Many Sweeps**

**Step 2.3.1:** Create Combination for Texas's Sweep 1: SSN4 + Last Name + DOB.

**Step 2.3.2:** Create Combination for Texas's Sweep 3: First Name + Last Name + DOB

**Step 2.3.3:** Create Combination for Texas's Sweep 4: First Name + Middle Initial + Last Name + DOB

**Note:** Combinations for Texas's Sweeps 1 and 3 do not already exist as pre-made fields in the TEAM database extract but instead must be created from the underlying TEAM database fields, in addition to being constructed on the Federal database side. Texas's Sweep 4 is equivalent to

---

<sup>4</sup> Only for the State Department, create three further variations of Combination G created using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

Exhibit A

the combination for the United States' Combination G. Texas's Sweep 2 is on full 9 social security number.<sup>5</sup>

**Stage 4: Establish Identifier Uniqueness For Combinations A - L**

**Step 2.4.1:** Generate a field that establishes the uniqueness of each identifier variable.

For federal databases, for each combination A-L, generate a field that establishes uniqueness among only Texas records and a field that establishes uniqueness among nationwide records. *E.g.*, if only one record has the string <JEANSMITH01011950012377777> for Combination A, populate the uniqueness field for Combination A for that record as <1>. If four records have the string <JOHNSMITHA0101950> for Combination G, populate the uniqueness field for Combination G for each of those records as <2>, which indicates any number greater than one.

---

<sup>5</sup> For purposes of matching to U.S. Department of State Passport and Passport Card holder data, Texas's Sweep 3 and Sweep 4 are as follows: Sweep 3: Last Name + First Name (restricted to 32 characters) + DOB. Sweep 4: Last Name + First Name + Middle Initial (restricted to 32 characters) + DOB.

Exhibit A

### **PART III: MATCH DATABASES**

#### **Stage 1: United States' Primary One-to-Many Matching Sweeps**

- Step 3.1.1:** For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identifier or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).
- Step 3.1.2:** Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identifier or disability database.
- Step 3.1.3:** Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identifier databases produced by the State of Texas.

Exhibit A

**Stage 2: United States' Secondary One-to-Many Matching Sweeps**

- Step 3.2.1:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identifier or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).
- Step 3.2.2:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number<sup>6</sup> in the TEAM database against the equivalent combination/field in the identifier or disability database.
- Step 3.2.3:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identifier or disability database.

---

<sup>6</sup> The full social security number is not created as a separate “combination” as it is its own field stored within the TEAM database under the field name “ssn”.



Exhibit A

**Step 3.2.4:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identifier or disability database.

**Stage 3: United States' Nationwide Federal Sweeps**

**Step 3.3.1:** For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identifier or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

**Step 3.3.2:** For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identifier or disability database.

**Step 3.3.3:** For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination

## Exhibit A

K against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.

**Step 3.3.4:** For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination L against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.<sup>7</sup>

**Stage 4: Texas' Many-to-Many Nationwide Sweeps**

**Step 3.4.1** Regardless of whether the combination for Sweep 1 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

---

<sup>7</sup> Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last\_fix + First\_fix + Middle\_name from the TEAM database;
- Combination G2 to DOB + Last\_fix + First\_fix from the TEAM database;
- Combination G2 to DOB + Last\_fix + First word of First\_name from the TEAM database;
- Combination G3 to DOB + Last\_fix + First\_fix + Middle\_Initial from the TEAM database;
- Combination G3 to DOB + Last\_fix + First word of First\_name + Middle\_Initial from the TEAM database; and
- Combination G3 to DOB + Last\_fix + First word of First\_name + First character of Second word of First\_fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

Exhibit A

**Step 3.4.2** Regardless of whether full 9 social security number is unique in the TEAM database, for Sweep 2, match against the equivalent field in a nationwide search of the Federal database.

**Step 3.4.3** Regardless of whether the combination for Sweep 3 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

**Step 3.4.4** Regardless of whether the combination for Sweep 4 (Combination G) is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

**Note:** For each of the Texas many-to-many sweeps:

- Indicate <1> if any TEAM combination matches a single combination in the Federal database
- Indicate <2> if any TEAM combination matches more than one record in the Federal database.
- Indicate <0> if no match is achieved.

**Examples:**

- If there are two TEAM records that have identical versions of the combination for Sweep 1, and there is one record in the Federal database that matches on this combination, both of the underlying TEAM records will have a matching output of <1> for Sweep 1.
- If there are three TEAM records that have identical versions of the combination for Sweep 3, and there are five records in the Federal database that match on that combination, the three TEAM records will each have a matching output of <2> for Sweep 3.

## Exhibit B

**Step 1.3 Diagnostics on the VHA Data**

There were 5,020,207 records extracted from the VHA data for Veterans with a Veteran Identification Card (VIC) on 2/28/2014. The number of Veterans with a VIC and a Texas residential address was 389,383.

**Name**

	Number of Records	Percent of Records
Last Name Missing	0	0.00%
First Name Missing	31	0.00%
First Name Only One Character	11,881	0.24%
Middle Initial Missing	661,201	13.17%

**Social Security Number**

	Number of Records	Percent of Records
SSN Missing	2,481	0.05%
Invalid SSN (begins with '00000')	27	0.00%

**Gender**

Gender was 100% populated.

**Date of Birth**

The possibly invalid dates of birth below appear to be default values used in the date of birth (e.g., first day of month, 15<sup>th</sup> of month, first day of year). This is rare.

	Number of Records	Percent of Records
Incomplete date or outside 1901-1996	77	0.00%
Possibly invalid (01-01-01, etc.)	21	0.00%

**Zip Code**

	Number of Records	Percent of Records
Zip missing (Most of these were cases with a missing residential address)	132,103	2.63%

Exhibit B

### **Mailing Zip**

There were 5,301 veterans with a mailing address.

	Number of Records	Percent of Records
Mailing Zip missing	20	0.38%

IN THE UNITED STATES DISTRICT COURT  
FOR THE SOUTHERN DISTRICT OF TEXAS  
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*,

Plaintiffs,

v.

RICK PERRY, *et al.*,

Defendants.

Civil Action No. 2:13-cv-193 (NGR)

UNITED STATES OF AMERICA,

Plaintiff,

TEXAS LEAGUE OF YOUNG VOTERS  
EDUCATION FUND, *et al.*,

Plaintiff-Intervenors,

TEXAS ASSOCIATION OF HISPANIC  
COUNTY JUDGES AND COUNTY  
COMMISSIONERS, *et al.*,

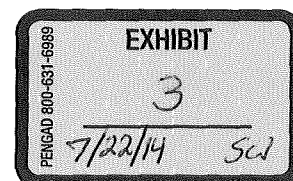
Plaintiff-Intervenors,

v.

STATE OF TEXAS, *et al.*,

Defendants.

Civil Action No. 2:13-cv-263 (NGR)



TEXAS STATE CONFERENCE OF NAACP  
BRANCHES, *et al.*,

Plaintiffs,

v.

JOHN STEEN, *et al.*,

Defendants.

Civil Action No. 2:13-cv-291 (NGR)

BELINDA ORTIZ, *et al.*,

Plaintiffs,

v.

STATE OF TEXAS, *et al.*,

Defendants

Civil Action No. 2:13-cv-348 (NGR)

**DECLARATION OF YEMI B. OSHINNAIYE**

I, Yemi B. Oshinnaiye, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as Acting Associate Chief of Systems Operations, Office of Information Technology ("OIT"), U.S. Citizenship and Immigration Services ("USCIS"), Department of Homeland Security. I have served in this capacity since May 15, 2014. My duties include managing Branch and Project Managers that oversee development, deployment and production support for systems that support USCIS current operations. Prior to this position I served as Branch Chief for Benefits and Biometrics, USCIS OIT. I served in that position beginning in October, 2012. My duties included Management of Project/Program managers that oversaw the development, deployment and production support for systems that supported the administration, management and granting of



benefits from USCIS and systems that validate eligibility for those benefits. I have personal knowledge of the information contained in this declaration based upon my work for the USCIS, and my personal involvement in and supervision of the database comparison described below.

2. On February 27, 2014, my agency received an encrypted hard drive containing data regarding Texas registered voters from the United States Department of Justice ("DOJ") in the form of a 7 gigabyte .csv file containing 13,564,420 records for Texas registered voters. USCIS was able to load all 13,564,420 records.

3. The Texas data that USCIS received contained the following fields:

- Vuid	- str_num_res_alt
- last_name	- AGDN
- first_name	- AGDN_TAG
- middle_name	- AGDN_unique
- ssn	- AGDNlast
- ssn4_str	- AGDNlast_TAG
- gender_code	- AGDNlast_unique
- official_id	- AGD
- dl_id_str	- AGD_TAG
- date_of_birth	- AGD_unique
- voter_status_code	- ADN
- block_number	- ADN_TAG
- street_name	- ADN_unique
- city	- AGN
- state	- AGN_TAG
- zip_code	- AGN_unique
- zip_str	- GDN
- mail_adrs_1	- GDN_TAG
- last_fix	- GDN_unique
- multi_lastname1	- DLN
- multi_lastname2	- DLN_TAG
- first_fix	- DLN_unique
- middle_initial	- unique_count
- fern_str	- FMLD
- dob_str	- FMLD_TAG
- zip5_res_alt	- FMLD_unique

- |                |                 |
|----------------|-----------------|
| - SDZ          | - FML2D         |
| - SDZ_TAG      | - FML2D_TAG     |
| - SDZ_unique   | - FML2D_unique  |
| - SND          | - unique_count2 |
| - SND_TAG      | - var_fill      |
| - SND_unique   | - K2            |
| - FML1D        | - K3            |
| - FML1D_TAG    | - L2            |
| - FML1D_unique | - L3            |

4. USCIS executed the data preparation and comparison algorithm as it was relevant to USCIS databases containing U.S. Certificate of Citizenship or U.S. Certificate of Naturalization data , attached as Ex. A, that was provided by DOJ.<sup>1</sup>
5. Through this process, USCIS attempted to match particular combinations of identifying information for Texas registered voters (*e.g.*, first and last name, gender, and date of birth) with the same or related combinations of identifying information with respect to holders of a U.S. Certificate of Citizenship or a U.S. Certificate of Naturalization found in the USCIS Computer Linked Application Information Management System 4 (“CLAIMS 4”) and Central Index System (“CIS”), which contain U.S. Certificate of Citizenship or U.S. Certificate of Naturalization data. For each particular combination, USCIS identified instances where the identifying information from the Texas registered voter information matched with the same or related combinations of identifying information for one or more records of individuals issued a U.S. Certificate of Citizenship or a U.S. Certificate of Naturalization. The frequencies of any missing values in the underlying USCIS data are provided in Ex. B.
6. For each sweep through the relevant USCIS dataset, we tracked the results,

---

<sup>1</sup> The only data preparation step listed in Ex. A that USCIS was unable to complete was the Step 1.7.2 imputation of gender for USCIS records with addresses outside of Texas. No gender imputation was conducted for USCIS records with an address outside of Texas that contained a blank gender value.

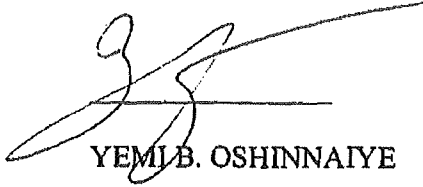
on a record-by-record basis, of the database comparisons on all of the requested combinations. In order to speed computer processing time, the results were appended only to a unique identifier, the Var\_Fill value, that had been provided for each Texas voter registration record we received. No other underlying data from the voter registration database was included in the results file sent to DOJ.

7. On May 23, 2014, USCIS transferred responsive data to DOJ, in the form of a 535 megabyte text (.txt) file, with comma delimiters.
8. The results of each sweep, as described in Ex. A, can be found in the following columns of the data USCIS returned to DOJ:
  - a. "USA" contains results of the Stage 1, Step 3.1.1, Combination A match;
  - b. "USB" contains results of the Stage 1, Step 3.1.2, Combination B match;
  - c. "USC" contains results of the Stage 1, Step 3.1.2, Combination C match;
  - d. "USD" contains results of the Stage 1, Step 3.1.2, Combination D match;
  - e. "USE" contains results of the Stage 1, Step 3.1.2, Combination E match;
  - f. "USF" contains results of the Stage 1, Step 3.1.2, Combination F match;
  - g. "USG" contains results of the Stage 2, Step 3.2.1, Combination G match;
  - h. "USH" contains results of the Stage 2, Step 3.2.2, Combination H match;
  - i. "USI" contains results of the Stage 2, Step 3.2.2, Combination I match;
  - j. "USSSN" contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;
  - k. "USGK" contains results of the Stage 2, Step 3.2.3, Combination K to G match;
  - l. "USK" contains results of the Stage 2, Step 3.2.3, Combination K match;
  - m. "USKL" contains results of the Stage 2, Step 3.2.3, Combination K to L match;

- n. "USGL" contains results of the Stage 2, Step 3.2.4, Combination L to G match;
  - o. "USL" contains results of the Stage 2, Step 3.2.4, Combination L match;
  - p. "USLK" contains results of the Stage 2, Step 3.2.4, Combination L to K match;
  - q. "USF-NW" contains results of the Stage 3, Step 3.3.1, Combination F match;
  - r. "USG-NW" contains results of the Stage 3, Step 3.3.2, Combination G match;
  - s. "USI-NW" contains results of the Stage 3, Step 3.3.2, Combination I match;
  - t. "USSSN-NW" contains results of the Stage 3, Step 3.3.2, nine-digit social security number match;
  - u. "USGK\_NW" contains results of the Stage 3, Step 3.3.3, Combination K to G match;
  - v. "USK\_NW" contains results of the Stage 2, Step 3.3.3, Combination K match;
  - w. "USKL\_NW" contains results of the Stage 2, Step 3.3.3, Combination K to L match;
  - x. "USGL\_NW" contains results of the Stage 3, Step 3.3.4, Combination L to G match;
  - y. "USL\_NW" contains results of the Stage 3, Step 3.3.4, Combination L match;
  - z. "USLK\_NW" contains results of the Stage 3, Step 3.3.4, Combination L to K match;
  - aa. "Texas1" contains results of the Stage 4, Step 3.4.1, Sweep 1 match;
  - bb. "Texas2" contains results of the Stage 4, Step 3.4.2, Sweep 2 match;
  - cc. "Texas3" contains results of the Stage 4, Step 3.4.3, Sweep 3 match; and
  - dd. "Texas4" contains results of the Stage 4, Step 3.4.4, Sweep 4 match.
9. No other matching comparisons between the Texas data and data from USCIS were undertaken beyond those set forth in Ex. A.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on May 29, 2014.

A handwritten signature in black ink, consisting of stylized, overlapping loops and a long horizontal stroke extending to the right, positioned above a solid horizontal line.

YEMI B. OSHINNAIYE

Acting Associate Chief of Systems  
Operations, Office of Information  
Technology, U.S. Citizenship and  
Immigration Services





Exhibit A

**United States v. Texas: Federal Agency Algorithm Summary**

This document summarizes the database matches that both the United States and all other plaintiffs and the State of Texas have requested from Federal Agencies as part of the *Veasey v. Perry/United States v. Texas* litigation (S.D. Tex).

The matching process proceeds in three parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, the United States' one-to-many matches and the State of Texas's many-to-many matches are conducted between databases.

**PART I: DATABASE PREPARATION**

**Stage 1: Extraction of Available Data from Federal Identification & Disability Databases**

- Step 1.1.1:** Extract complete name into separate first name, middle name, and last name fields.
- Step 1.1.2:** Extract date of birth.
- Step 1.1.3:** Extract gender.
- Step 1.1.4:** Extract residential address and mailing address.
- Step 1.1.5:** Extract social security number.
- Step 1.1.6:** Extract Texas driver license number (only if present in Federal database).

**Stage 2: Separate Valid Identification and Disability Records**

- Step 1.2.1:** Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

Exhibit A

- Step 1.2.2:** Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

**Stage 3: Diagnostics**

- Step 1.3.1:** Report the frequency of missing values for each field.
- Step 1.3.2:** Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.
- Step 1.3.3:** Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

**Stage 4: Standardize Last Name**

- Step 1.4.1:** Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.
- Step 1.4.2:** For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.
- Step 1.4.3:** Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.
- Step 1.4.4:** Code all missing values as blank fields.

Exhibit A

**Stage 5: Standardize First Name and Middle Name**

**Step 1.5.1:** Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.

**Step 1.5.2:** Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.<sup>1</sup>

**Step 1.5.3:** Code all missing values as blank fields.

**Stage 6: Standardize Date of Birth**

**Step 1.6.1:** Convert the date of birth to an eight-digit string of MMDDYYYY.

**Step 1.6.2:** Code all missing values as blank fields.

**Stage 7: Standardize Gender**

**Step 1.7.1:** Code gender as a string of 1 for females and 0 for males.

**Step 1.7.2:** Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.

**Step 1.7.3:** If missing values remain, code all missing values as blank fields.

---

<sup>1</sup> The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Exhibit A

**Stage 8: Standardize Address**

- Step 1.8.1:** Convert the residential ZIP code to a string if it is stored as a numeric field.
- Step 1.8.2:** Where the residential address ZIP code is blank, populate that field with the value in the mailing address ZIP code field, if available.<sup>2</sup>
- Step 1.8.3:** Truncate the residential ZIP code field to the first five digits. *E.g.*, <77777-1234> becomes <77777>.
- Step 1.8.4:** Where the residential address field is blank, populate that field with the value in the mailing address field, if available.
- Step 1.8.5:** Where address field containing street address begins with a street number, isolate the street number. *E.g.*, <123 Main Street> becomes <123>.
- Step 1.8.6:** Where the address field begins with recognized strings indicating a mail box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444> becomes <444>.
- Step 1.8.7:** If missing values remain, code all missing values as blank fields.

**Stage 9: Standardize Social Security Number**

- Step 1.9.1:** Convert the social security number to a string if it is stored as a numeric field.
- Step 1.9.2:** Using full social security number, check for invalid SSNs. In the case of invalid SSNs, code as missing. *E.g.*, <123456789> becomes <◇>.

---

<sup>2</sup> For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

Exhibit A

**Step 1.9.3:** Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

**Step 1.9.4:** Code all missing values as blank fields.

**PART II: DATABASE PREPARATION**

**Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps**

**Step 2.1.1:** Create Combination A: First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.<sup>3</sup>

**Step 2.1.2:** Create Combination B: Last Name + Gender + DOB + Residential ZIP + Residential Street Number.

**Step 2.1.3:** Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.

**Step 2.1.4:** Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.

**Step 2.1.5:** Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.

**Step 2.1.6:** Create Combination F: First Name + Last Name + Gender + DOB.

**Step 2.1.7:** Create Combination M: Texas Driver License Number (where available).

---

<sup>3</sup> For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Exhibit A

**Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps**

- Step 2.2.1:** Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.<sup>4</sup>
- Step 2.2.2:** Create Combination H: SSN4 + Date of Birth + Residential ZIP.
- Step 2.2.3:** Create Combination I: SSN4 + First Name + Last Name + Date of Birth.
- Step 2.2.4:** Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.
- Step 2.2.5:** Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.
- Step 2.2.6:** Full Social Security Number.

**Stage 3: Construct Identifiers Used Only For Texas's Many-to-Many Sweeps**

- Step 2.3.1:** Create Combination for Texas's Sweep 1: SSN4 + Last Name + DOB.
- Step 2.3.2:** Create Combination for Texas's Sweep 3: First Name + Last Name + DOB
- Step 2.3.3:** Create Combination for Texas's Sweep 4: First Name + Middle Initial + Last Name + DOB

**Note:** Combinations for Texas's Sweeps 1 and 3 do not already exist as pre-made fields in the TEAM database extract but instead must be created from the underlying TEAM database fields, in addition to being constructed on the Federal database side. Texas's Sweep 4 is equivalent to

---

<sup>4</sup> Only for the State Department, create three further variations of Combination G created using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

Exhibit A

the combination for the United States' Combination G. Texas's Sweep 2 is on full 9 social security number.<sup>5</sup>

**Stage 4: Establish Identifier Uniqueness For Combinations A - L**

**Step 2.4.1:** Generate a field that establishes the uniqueness of each identifier variable.

For federal databases, for each combination A-L, generate a field that establishes uniqueness among only Texas records and a field that establishes uniqueness among nationwide records. *E.g.*, if only one record has the string <JEANSMITH01011950012377777> for Combination A, populate the uniqueness field for Combination A for that record as <1>. If four records have the string <JOHNSMITHA0101950> for Combination G, populate the uniqueness field for Combination G for each of those records as <2>, which indicates any number greater than one.

---

<sup>5</sup> For purposes of matching to U.S. Department of State Passport and Passport Card holder data, Texas's Sweep 3 and Sweep 4 are as follows: Sweep 3: Last Name + First Name (restricted to 32 characters) + DOB. Sweep 4: Last Name + First Name + Middle Initial (restricted to 32 characters) + DOB.



Exhibit A

### **PART III: MATCH DATABASES**

#### **Stage 1: United States' Primary One-to-Many Matching Sweeps**

- Step 3.1.1:** For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identifier or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).
- Step 3.1.2:** Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identifier or disability database.
- Step 3.1.3:** Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identifier databases produced by the State of Texas.

Exhibit A

**Stage 2: United States' Secondary One-to-Many Matching Sweeps**

**Step 3.2.1:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identifier or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

**Step 3.2.2:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number<sup>6</sup> in the TEAM database against the equivalent combination/field in the identifier or disability database.

**Step 3.2.3:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identifier or disability database.

---

<sup>6</sup> The full social security number is not created as a separate "combination" as it is its own field stored within the TEAM database under the field name "ssn".

Exhibit A

**Step 3.2.4:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identifier or disability database.

**Stage 3: United States' Nationwide Federal Sweeps**

**Step 3.3.1:** For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identifier or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

**Step 3.3.2:** For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identifier or disability database.

**Step 3.3.3:** For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination

Exhibit A

K against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.

**Step 3.3.4:** For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination L against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.<sup>7</sup>

**Stage 4: Texas' Many-to-Many Nationwide Sweeps**

**Step 3.4.1** Regardless of whether the combination for Sweep 1 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

---

<sup>7</sup> Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last\_fix + First\_fix + Middle\_name from the TEAM database;
- Combination G2 to DOB + Last\_fix + First\_fix from the TEAM database;
- Combination G2 to DOB + Last\_fix + First word of First\_name from the TEAM database;
- Combination G3 to DOB + Last\_fix + First\_fix + Middle\_Initial from the TEAM database;
- Combination G3 to DOB + Last\_fix + First word of First\_name + Middle\_Initial from the TEAM database; and
- Combination G3 to DOB + Last\_fix + First word of First\_name + First character of Second word of First\_fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

Exhibit A

**Step 3.4.2** Regardless of whether full 9 social security number is unique in the TEAM database, for Sweep 2, match against the equivalent field in a nationwide search of the Federal database.

**Step 3.4.3** Regardless of whether the combination for Sweep 3 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

**Step 3.4.4** Regardless of whether the combination for Sweep 4 (Combination G) is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

**Note:** For each of the Texas many-to-many sweeps:

- Indicate <1> if any TEAM combination matches a single combination in the Federal database
- Indicate <2> if any TEAM combination matches more than one record in the Federal database.
- Indicate <0> if no match is achieved.

**Examples:**

- If there are two TEAM records that have identical versions of the combination for Sweep 1, and there is one record in the Federal database that matches on this combination, both of the underlying TEAM records will have a matching output of <1> for Sweep 1.
- If there are three TEAM records that have identical versions of the combination for Sweep 3, and there are five records in the Federal database that match on that combination, the three TEAM records will each have a matching output of <2> for Sweep 3.

Exhibit B

**United States Citizenship and Immigration Services****Step 1.3 Diagnostics**

USCIS had a population of 19,401,814 United States Citizens who were presented either the N550 Naturalization Certificate or the N560 Certificate of Citizenship. This listing was extracted from the USCIS Central Index System which is the most comprehensive listing of United States Citizens.

Below are the numbers of USCIS records analyzed missing information from a particular listed field.

**Step 1.3.1**

Missing first name – 49,814  
 Missing middle name – 7,180,037  
 Missing last name – 0  
 Missing date of birth – 18,632  
 Missing gender – 8,251,075  
 Missing SSN – 14,264,377

**Step 1.3.2:****Number of Records with Invalid SSNs:**

5	123456789
2	987654321
12	111111111
3	555555555
2	666666666
1	777777777
2	888888888
40	999999999

**Step 1.3.3****Number of Records with Potentially Invalid dates of birth**

85	19010101
49	19020202
29	19030303
41	19040404
59	19050505
44	19060606
55	19070707
58	19080808
69	19090909
103	19111111
122	19121212

IN THE UNITED STATES DISTRICT COURT  
FOR THE SOUTHERN DISTRICT OF TEXAS  
CORPUS CHRISTI DIVISION

MARC VEASEY, *et al.*,

Plaintiffs,

v.

RICK PERRY, *et al.*,

Defendants.

Civil Action No. 2:13-cv-193 (NGR)

UNITED STATES OF AMERICA,

Plaintiff,

TEXAS LEAGUE OF YOUNG VOTERS  
EDUCATION FUND, *et al.*,

Plaintiff-Intervenors,

TEXAS ASSOCIATION OF HISPANIC  
COUNTY JUDGES AND COUNTY  
COMMISSIONERS, *et al.*,

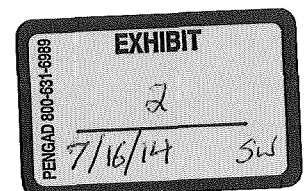
Plaintiff-Intervenors,

v.

STATE OF TEXAS, *et al.*,

Defendants.

Civil Action No. 2:13-cv-263 (NGR)





TEXAS STATE CONFERENCE OF NAACP  
BRANCHES, *et al.*,

Plaintiffs,

v.

JOHN STEEN, *et al.*,

Defendants.

Civil Action No. 2:13-cv-291 (NGR)

BELINDA ORTIZ, *et al.*,

Plaintiffs,

v.

STATE OF TEXAS, *et al.*,

Defendants

Civil Action No. 2:13-cv-348 (NGR)

**DECLARATION OF Michelle Saunders Rudolph**

I, Michelle Saunders Rudolph, pursuant to 28 U.S.C. § 1746, declare as follows:

1. I currently serve as Acting Division Director, Data, Analysis and Programs Division, DMDC/DHRA/DoD. I have served in this capacity for close to 2 years. My duties include ensuring the data arriving at DMDC is of sound quality so Servicemembers and their families can receive the benefits and entitlements they have earned. In addition, I ensure agencies and offices with a need to know, receive accurate and timely data.
2. I have personal knowledge of the information contained in this declaration based upon my work for the Defense Manpower Data Center at the U.S. Department of Defense ("DMDC"), and my supervision of completion of the database comparison described below.

3. On February 28, 2014, DMDC received from the United States Department of Justice an encrypted hard drive containing data regarding Texas registered voters in the form of a 7 gigabyte .csv file containing 13,564,420 records for Texas registered voters. DMDC was able to load all 13,564,420 records.

4. The Texas data that DMDC received contained the following fields:

- Vuid	- AGD
- last_name	- AGD_TAG
- first_name	- AGD_unique
- middle_name	- and
- ssn	- ADN_TAG
- ssn4_str	- ADN_unique
- gender_code	- AGN
- official_id	- AGN_TAG
- dl_id_str	- AGN_unique
- date_of_birth	- GDN
- voter_status_code	- GDN_TAG
- block_number	- GDN_unique
- street_name	- DLN
- city	- DLN_TAG
- state	- DLN_unique
- zip_code	- unique_count
- zip_str	- FMLD
- mail_adrs_1	- FMLD_TAG
- last_fix	- FMLD_unique
- multi_lastname1	- SDZ
- multi_lastname2	- SDZ_TAG
- first_fix	- SDZ_unique
- middle_initial	- SND
- fem_str	- SND_TAG
- dob_str	- SND_unique
- zip5_res_alt	- FML1D
- str_num_res_alt	- FML1D_TAG
- AGDN	- FML1D_unique
- AGDN_TAG	- FML2D
- AGDN_unique	- FML2D_TAG
- AGDNlast	- FML2D_unique
- AGDNlast_TAG	- unique_count2
- AGDNlast_unique	- var_fill

- K2
- K3
- L2
- L3

5. DMDC executed the data preparation and comparison steps, attached as Ex. A, that were provided by the United States Department of Justice.
6. Through this process, DMDC attempted to match particular combinations of identifying information for Texas registered voters (*e.g.*, first and last name, gender, and date of birth) with the same or related combinations of identifying information with respect to persons eighteen years of age or older holding a valid form of one of the following Department of Defense (“DoD”) photo identification cards: Common Access Cards and Uniformed Services ID Cards (including Department of Defense Civilian Retiree Cards). For each particular combination, DMDC identified instances where the identifying information for a Texas registered voter matched with the respective combination of identifying information for one or more holders of DoD photo identification.
7. In order to conduct the requested matches, and to create and record the combinations used for the Stage 4 matches, DMDC reformatted the original .csv file into a fixed format flat file to which matching results were appended.
8. Two DoD datasets were created for this analysis—one contained records of holders of qualifying DoD photo ID located in Texas and the other contained records for holders of qualifying DoD photo ID located outside of Texas. For each sweep through the relevant DoD datasets, we appended a column to the Texas data to indicate, on a record-by-record basis, the output of the database comparisons on all of the requested combinations. The frequencies of any missing values in the underlying DoD data are provided as Ex. B.

9. On May 9, 2014, DMDC transferred responsive data to the Department of Justice, in the form of a 21.5 gigabyte file with pipe delimiters. The name of the results file transferred to the Department of Justice is "Team Match Report".
10. The results of each sweep, as described in Ex. A, can be found in the following columns of the data that DMDC returned to the Department of Justice on May 9:
  - a. "Key\_A\_match\_flag\_Texas" contains results of the Stage 1, Step 3.1.1, Combination A match;
  - b. "Key\_B\_match\_flag\_Texas" contains results of the Stage 1, Step 3.1.2, Combination B match;
  - c. "Key\_C\_match\_flag\_Texas" contains results of the Stage 1, Step 3.1.2, Combination C match;
  - d. "Key\_D\_match\_flag\_Texas" contains results of the Stage 1, Step 3.1.2, Combination D match;
  - e. "Key\_E\_match\_flag\_Texas" contains results of the Stage 1, Step 3.1.2, Combination E match;
  - f. "Key\_F\_match\_flag\_Texas" contains results of the Stage 1, Step 3.1.2, Combination F match;
  - g. "Key\_G\_match\_flag\_Texas" contains results of the Stage 2, Step 3.2.1, Combination G match;
  - h. "Key\_H\_match\_flag\_Texas" contains results of the Stage 2, Step 3.2.2, Combination H match;
  - i. "Key\_I\_match\_flag\_Texas" contains results of the Stage 2, Step 3.2.2, Combination I match;

- j. "Key\_KtoG\_match\_flag\_Texas" contains results of the Stage 2, Step 3.2.3,  
Combination K to G match;
- k. "Key\_KtoK\_match\_flag\_Texas" contains results of the Stage 2, Step 3.2.3,  
Combination K match;
- l. "Key\_KtoL\_match\_flag\_Texas" contains results of the Stage 2, Step 3.2.3,  
Combination K to L match;
- m. "Key\_LtoG\_match\_flag\_Texas" contains results of the Stage 2, Step 3.2.4,  
Combination L to G match;
- n. "Key\_LtoL\_match\_flag\_Texas" contains results of the Stage 2, Step 3.2.4,  
Combination L match;
- o. "Key\_LtoK\_match\_flag\_Texas" contains results of the Stage 2, Step 3.2.4,  
Combination L to K match;
- p. "Key\_F\_match\_flag\_Nationwide" contains results of the Stage 3, Step 3.3.1,  
Combination F match;
- q. "Key\_I\_match\_flag\_Nationwide" contains results of the Stage 3, Step 3.3.2,  
Combination I match;
- r. "Key\_KtoG\_match\_flag\_Nationwide" contains results of the Stage 3, Step 3.3.3,  
Combination K to G match;
- s. "Key\_KtoK\_match\_flag\_Nationwide" contains results of the Stage 2, Step 3.3.3,  
Combination K match;
- t. "Key\_KtoL\_match\_flag\_Nationwide" contains results of the Stage 2, Step 3.3.3,  
Combination K to L match;

- u. "Key\_LtoG\_match\_flag\_Nationwide" contains results of the Stage 3, Step 3.3.4, Combination L to G match;
- v. "Key\_LtoL\_match\_flag\_Nationwide" contains results of the Stage 3, Step 3.3.4, Combination L match;
- w. "Key\_LtoK\_match\_flag\_Nationwide" contains results of the Stage 3, Step 3.3.4, Combination L to K match;

11. In the May 9 dataset provided to the Department of Justice, no results were provided for the following matches:

- a. the Stage 2, Step 3.2.2, nine-digit social security number match;
- b. the Stage 3, Step 3.3.2, Combination G match; and
- c. the Stage 3, Step 3.3.2, nine-digit social security number match.

These match results were provided separately on to the Department of Justice on May 28, 2014 as described in paragraph 13 below.

12. In the original May 9 dataset provided to the Department of Justice, results for the Stage 4 matches were provided in the following fields:

- a. "Sweep\_1\_match\_flag\_Nationwide" had partial results for the Stage 4, Step 3.4.1, Sweep 1 match;
- b. "Sweep\_2\_match\_flag\_Nationwide" had partial results for the Stage 4, Step 3.4.2, Sweep 2 match;
- c. "Sweep\_3\_match\_flag\_Nationwide" had partial results for the Sweep 3 match; and
- d. "Sweep\_4\_match\_flag\_Nationwide" had partial results for the Stage 4, Step 3.4.4, Sweep 4 match.

For these Stage 4 matches in the May 9 dataset, the only voter registration records compared were those that had not already matched on a Stage 1 through 3 sweep. Moreover, the only DoD records compared were for qualifying ID holders outside of Texas. At the request of the Department of Justice, DMDC re-ran each of the Stage 4 matches, matching all Texas voter registration records regardless of whether they had been matched on a Stage 1 through 3 sweep. In addition, the complete voter registration database extract was matched against all qualifying DoD records for ID holders in and outside of Texas. These corrected and complete match results for the Stage 4 sweeps were provided to Department of Justice as part of the May 28, 2014 dataset.

13. The May 28 dataset thus supplemented data not previously provided by DMDC as to three of the sweeps requested by the United States, and also re-ran each of the Stage 4 matches requested by Defendants. To enable the results from the May 28 dataset to be integrated with the May 9 dataset, the May 28 dataset included two unique identifiers: the VUID field from the voter registration file, which is almost completely unique, as well as the Var\_Fill field, which is completely unique. In addition to these identifiers, the following results fields were included in the May 28 dataset:
  - a. "SWEEP\_1\_MATCH\_FLAG\_TEXAS" contains results of the Stage 4, Step 3.4.1, Sweep 1 match based on the Texas-only dataset and  
"SWEEP\_1\_MATCH\_FLAG\_NATIONWIDE" contains the results for the same step based on the non-Texas, nationwide dataset;
  - b. "SWEEP\_2\_MATCH\_FLAG\_TEXAS" contains results of the Stage 4, Step 3.4.2, Sweep 2 match based on the Texas-only dataset and

- “SWEEP\_2\_MATCH\_FLAG\_NATIONWIDE” contains the results for the same step based on the non-Texas, nationwide dataset;
- c. “SWEEP\_3\_MATCH\_FLAG\_TEXAS” contains results of the Stage 4, Step 3.4.3, Sweep 3 match based on the Texas-only dataset and “SWEEP\_3\_MATCH\_FLAG\_NATIONWIDE” contains the results for the same step based on the non-Texas, nationwide dataset;
- d. “SWEEP\_4\_MATCH\_FLAG\_TEXAS” contains results of the Stage 4, Step 3.4.4, Sweep 4 match based on the Texas-only dataset and “SWEEP\_4\_MATCH\_FLAG\_NATIONWIDE” contains the results for the same step based on the non-Texas, nationwide dataset;
- e. “SSN9\_SECONDARY\_TEXAS” contains results of the Stage 2, Step 3.2.2, nine-digit social security number match;
- f. “SSN9\_SECONDARY\_NATIONWIDE” contains the results of the Stage 3, Step 3.3.2, nine-digit social security number match; and
- g. “KEY\_G\_MATCH\_FLAG\_NATIONWIDE” contains the results of the Stage 3, Step 3.3.2, Combination G match.
14. In both the May 9 and May 28 datasets, a result of “1” indicates that the respective combination from one DoD record matched; “2” indicates that two or more DoD records matched on the respective combination; “0” indicates that there was no match; and a blank space indicates that no match was attempted.<sup>1</sup>

---

<sup>1</sup> The only exception to this convention is with respect to the Combination G match in the May 28 dataset, where a “0” can indicate both that no match was attempted and that no match was made. No match would have been attempted where the underlying voter registration record had matched on a Stage 1 or Stage 2 combination, or where the particular values in Combination G were not unique among all voter registration records to be compared.



15. No other matching comparisons between the Texas data and data from DMDC were undertaken beyond those set forth above, and in accordance with the steps provided in Ex. A.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on May 29, 2014.

A handwritten signature in black ink, appearing to read "Michelle Rudolph", with a long horizontal flourish extending to the right.

Exhibit A

**United States v. Texas: Federal Agency Algorithm Summary**

This document summarizes the database matches that both the United States and all other plaintiffs and the State of Texas have requested from Federal Agencies as part of the *Veasey v. Perry/United States v. Texas* litigation (S.D. Tex).

The matching process proceeds in three parts, which are explained in detail below. *First*, databases are prepared and standardized. *Second*, identifier values are constructed by combining multiple individual fields. *Third*, the United States' one-to-many matches and the State of Texas's many-to-many matches are conducted between databases.

**PART I: DATABASE PREPARATION**

**Stage 1: Extraction of Available Data from Federal Identification & Disability Databases**

- Step 1.1.1:** Extract complete name into separate first name, middle name, and last name fields.
- Step 1.1.2:** Extract date of birth.
- Step 1.1.3:** Extract gender.
- Step 1.1.4:** Extract residential address and mailing address.
- Step 1.1.5:** Extract social security number.
- Step 1.1.6:** Extract Texas driver license number (only if present in Federal database).

**Stage 2: Separate Valid Identification and Disability Records**

- Step 1.2.1:** Remove records from identification database extracts that indicate that an ID has been revoked or has expired more than 60 days before the date of the TEAM database snapshot (which is January 15, 2014).

Exhibit A

- Step 1.2.2:** Remove records from disability database extracts that do not indicate current disability status or indicate a Veterans Administration disability rating of less than 50%.

**Stage 3: Diagnostics**

- Step 1.3.1:** Report the frequency of missing values for each field.
- Step 1.3.2:** Report the frequencies of invalid Social Security numbers, such as 111111111 and 123456789.
- Step 1.3.3:** Report the frequencies of likely invalid dates of birth, such as January 1, 1901 and November 11, 1911.

**Stage 4: Standardize Last Name**

- Step 1.4.1:** Remove last name suffixes that are contained within the last name field, rather than a distinct suffix field. *E.g.*, <Smith Jr.> becomes <Smith>.
- Step 1.4.2:** For last names containing hyphens, populate separate last name fields for all parts of the last name. *E.g.*, the last name <Smith-Jones> would have the value <Smith> entered into a LastName1 field and the value <Jones> entered into a LastName2 field.
- Step 1.4.3:** Remove spaces, hyphens, periods, and apostrophes from all last name fields and convert all letters to uppercase. *E.g.*, <O'Connor> becomes <OCONNOR> and <Smith-Jones> becomes <SMITHJONES>.
- Step 1.4.4:** Code all missing values as blank fields.

Exhibit A

**Stage 5: Standardize First Name and Middle Name**

- Step 1.5.1:** Remove spaces, hyphens, periods, and apostrophes from the first name field and convert all letters to uppercase. *E.g.*, <Jean-Paul> becomes <JEANPAUL>.
- Step 1.5.2:** Parse the first letter of the middle name (if available) and use it to populate a middle initial field. *E.g.*, <John> would yield <J>.<sup>1</sup>
- Step 1.5.3:** Code all missing values as blank fields.

**Stage 6: Standardize Date of Birth**

- Step 1.6.1:** Convert the date of birth to an eight-digit string of MMDDYYYY.
- Step 1.6.2:** Code all missing values as blank fields.

**Stage 7: Standardize Gender**

- Step 1.7.1:** Code gender as a string of 1 for females and 0 for males.
- Step 1.7.2:** Fill missing gender values using the most common gender value for the first name associated with a record. *E.g.*, if 99% of records with first name <JOHN> are listed as male, assign the male identifier to all records with first name <JOHN> and no listed gender.
- Step 1.7.3:** If missing values remain, code all missing values as blank fields.

---

<sup>1</sup> The U.S. Department of State does not maintain a separate field for middle names in its database of U.S. Passport and Passport Card holders. Instead, both first and middle name may be stored in the first name field. For this database, the following rule will be applied: treat the first word in the first name field as the first name, and treat the first letter following the first space as the middle initial.

Exhibit A

**Stage 8: Standardize Address**

- Step 1.8.1:** Convert the residential ZIP code to a string if it is stored as a numeric field.
- Step 1.8.2:** Where the residential address ZIP code is blank, populate that field with the value in the mailing address ZIP code field, if available.<sup>2</sup>
- Step 1.8.3:** Truncate the residential ZIP code field to the first five digits. *E.g.*, <77777-1234> becomes <77777>.
- Step 1.8.4:** Where the residential address field is blank, populate that field with the value in the mailing address field, if available.
- Step 1.8.5:** Where address field containing street address begins with a street number, isolate the street number. *E.g.*, <123 Main Street> becomes <123>.
- Step 1.8.6:** Where the address field begins with recognized strings indicating a mail box, eliminate strings to isolate the box number. *E.g.*, <PO Box 444> becomes <444>.
- Step 1.8.7:** If missing values remain, code all missing values as blank fields.

**Stage 9: Standardize Social Security Number**

- Step 1.9.1:** Convert the social security number to a string if it is stored as a numeric field.
- Step 1.9.2:** Using full social security number, check for invalid SSNs. In the case of invalid SSNs, code as missing. *E.g.*, <123456789> becomes <>.

---

<sup>2</sup> For purposes of this database matching protocol, the only address fields utilized with respect to data regarding U.S. Passports and U.S. Passport Cards are those regarding mailing addresses.

## Exhibit A

**Step 1.9.3:** Extract the last four digits of full social security number as a four-character string and use them to populate a separate SSN4 field.

**Step 1.9.4:** Code all missing values as blank fields.

## **PART II: DATABASE PREPARATION**

### **Stage 1: Construct Primary Identifier Variables for United States' One-to-Many Sweeps**

- Step 2.1.1:** Create Combination A: First Name + Last Name + Gender + DOB + Residential ZIP + Residential Street Number. *E.g.*, the separate fields <JEAN>, <SMITH>, <0>, <01011950>, <77777>, and <123> are combined to a single field <JEANSMITH00101195077777123>.<sup>3</sup>
- Step 2.1.2:** Create Combination B: Last Name + Gender + DOB + Residential ZIP + Residential Street Number.
- Step 2.1.3:** Create Combination C: Gender + Date of Birth + Residential ZIP + Residential Street Number.
- Step 2.1.4:** Create Combination D: First Name + Last Name + Date of Birth + Residential ZIP + Residential Street Number.
- Step 2.1.5:** Create Combination E: First Name + Last Name + Gender + Residential ZIP + Residential Street Number.
- Step 2.1.6:** Create Combination F: First Name + Last Name + Gender + DOB.
- Step 2.1.7:** Create Combination M: Texas Driver License Number (where available).

---

<sup>3</sup> For the U.S. Department of State only, the name portion of any combination is truncated if it is more than 32 characters long.

Exhibit A

**Stage 2: Construct Secondary Identifier Variables for United States' One-to-Many Sweeps**

- Step 2.2.1:** Create Combination G: First Name + Middle Initial + Last Name + Date of Birth.<sup>4</sup>
- Step 2.2.2:** Create Combination H: SSN4 + Date of Birth + Residential ZIP.
- Step 2.2.3:** Create Combination I: SSN4 + First Name + Last Name + Date of Birth.
- Step 2.2.4:** Create Combination K: First Name + Last Name 1 + Middle Initial + Date of Birth.
- Step 2.2.5:** Create Combination L: First Name + Last Name 2 + Middle Initial + Date of Birth.
- Step 2.2.6:** Full Social Security Number.

**Stage 3: Construct Identifiers Used Only For Texas's Many-to-Many Sweeps**

- Step 2.3.1:** Create Combination for Texas's Sweep 1: SSN4 + Last Name + DOB.
- Step 2.3.2:** Create Combination for Texas's Sweep 3: First Name + Last Name + DOB
- Step 2.3.3:** Create Combination for Texas's Sweep 4: First Name + Middle Initial + Last Name + DOB

**Note:** Combinations for Texas's Sweeps 1 and 3 do not already exist as pre-made fields in the TEAM database extract but instead must be created from the underlying TEAM database fields, in addition to being constructed on the Federal database side. Texas's Sweep 4 is equivalent to

---

<sup>4</sup> Only for the State Department, create three further variations of Combination G created using the State Department's "LFMName" field which contains Last, First, and Middle Names, in that order, truncated to a maximum length of 32 characters. Combination G1 is DOB + LFMName; Combination G2 is DOB + First two words of LFMName; and Combination G3 is DOB + First two words of LFMName + First character of third word of LFMName.

Exhibit A

the combination for the United States' Combination G. Texas's Sweep 2 is on full 9 social security number.<sup>5</sup>

**Stage 4: Establish Identifier Uniqueness For Combinations A - L**

**Step 2.4.1:** Generate a field that establishes the uniqueness of each identifier variable.

For federal databases, for each combination A-L, generate a field that establishes uniqueness among only Texas records and a field that establishes uniqueness among nationwide records. *E.g.*, if only one record has the string <JEANSMITH0101195001237777> for Combination A, populate the uniqueness field for Combination A for that record as <1>. If four records have the string <JOHNSMITHA0101950> for Combination G, populate the uniqueness field for Combination G for each of those records as <2>, which indicates any number greater than one.

---

<sup>5</sup> For purposes of matching to U.S. Department of State Passport and Passport Card holder data, Texas's Sweep 3 and Sweep 4 are as follows: Sweep 3: Last Name + First Name (restricted to 32 characters) + DOB. Sweep 4: Last Name + First Name + Middle Initial (restricted to 32 characters) + DOB.



Exhibit A

### **PART III: MATCH DATABASES**

#### **Stage 1: United States' Primary One-to-Many Matching Sweeps**

- Step 3.1.1:** For each case in which Combination A is unique in the TEAM database, match Combination A against Combination A in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination A output field. Where there is a match, indicate the uniqueness of Combination A in the identifier or disability database in the Combination A output field (*e.g.*, in cases where there is one matching record in the Federal database, <1> should be inserted into the Combination A output field, while a <2> should be inserted into the Combination A output field if the TEAM record matched 2 or more records in the Federal database).
- Step 3.1.2:** Use the procedure in Step 3.1.1 to match Combination B, Combination C, Combination D, Combination E, and Combination F in the TEAM database against the equivalent combination field in the identifier or disability database.
- Step 3.1.3:** Use the procedure in Step 3.1.1 to match Combination M in the TEAM database against the equivalent combination field in the identifier databases produced by the State of Texas.

Exhibit A

**Stage 2: United States' Secondary One-to-Many Matching Sweeps**

**Step 3.2.1:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), and where Combination G is unique in the TEAM database, match Combination G against Combination G in the identifier or disability database. For federal databases, use only the subset of records with Texas addresses in the identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination G output field. Where there is a match, indicate the uniqueness of Combination G in the identifier or disability database in the Combination G output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

**Step 3.2.2:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination H, Combination I, and complete social security number<sup>6</sup> in the TEAM database against the equivalent combination/field in the identifier or disability database.

**Step 3.2.3:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination K against Combination G, Combination K, and Combination L in the identifier or disability database.

---

<sup>6</sup> The full social security number is not created as a separate "combination" as it is its own field stored within the TEAM database under the field name "ssn".

Exhibit A

**Step 3.2.4:** For each case in which no matches were found in the primary one-to-many matching sweeps (A-F, M), use the procedure in Step 3.2.1 to match Combination L against Combination G, Combination K, and Combination L in the identifier or disability database.

**Stage 3: United States' Nationwide Federal Sweeps**

**Step 3.3.1:** For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, and where Combination F is unique, match Combination F against Combination F in the nationwide identifier or disability database. Where a match is attempted but no match is found, indicate a zero in the Combination F nationwide output field. Where there is a match, indicate the uniqueness of Combination F in the identifier or disability database in the Combination F nationwide output field (*e.g.*, <1> if a unique match and <2> if matched to more than one record).

**Step 3.3.2:** For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination G, Combination I, and full social security number in the TEAM database against the equivalent combination/field in the nationwide identifier or disability database.

**Step 3.3.3:** For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination

## Exhibit A

K against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.

**Step 3.3.4:** For each case in which no matches were found in the primary and secondary matching sweeps of Texas records in a federal identifier or disability database, use the procedure in Step 3.3.1 to match Combination L against Combination G, Combination K, and Combination L in the nationwide identifier or disability database.<sup>7</sup>

**Stage 4: Texas' Many-to-Many Nationwide Sweeps**

**Step 3.4.1** Regardless of whether the combination for Sweep 1 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

---

<sup>7</sup> Step 3.3.5 for the State Department only: Match the following Combination G variations from applicable State Department records, first to include only the subset of records with Texas addresses, and then to include all applicable U.S. Passport and Passport Card records nationwide (*e.g.*, without Texas addresses), against the following fields from the TEAM database:

- Combination G1 to DOB + Last\_fix + First\_fix + Middle\_name from the TEAM database;
- Combination G2 to DOB + Last\_fix + First\_fix from the TEAM database;
- Combination G2 to DOB + Last\_fix + First word of First\_name from the TEAM database;
- Combination G3 to DOB + Last\_fix + First\_fix + Middle\_Initial from the TEAM database;
- Combination G3 to DOB + Last\_fix + First word of First\_name + Middle\_Initial from the TEAM database; and
- Combination G3 to DOB + Last\_fix + First word of First\_name + First character of Second word of First\_fix from the TEAM database.

Attempt matches for all TEAM records, regardless of whether they matched in any prior sweeps. Indicate <1> if a unique match and <2> if matched to more than one record.

Exhibit A

- Step 3.4.2** Regardless of whether full 9 social security number is unique in the TEAM database, for Sweep 2, match against the equivalent field in a nationwide search of the Federal database.
- Step 3.4.3** Regardless of whether the combination for Sweep 3 is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.
- Step 3.4.4** Regardless of whether the combination for Sweep 4 (Combination G) is unique in the TEAM database, match against the equivalent combination in a nationwide search of the Federal database.

**Note:** For each of the Texas many-to-many sweeps:

- Indicate <1> if any TEAM combination matches a single combination in the Federal database
- Indicate <2> if any TEAM combination matches more than one record in the Federal database.
- Indicate <0> if no match is achieved.

**Examples:**

- If there are two TEAM records that have identical versions of the combination for Sweep 1, and there is one record in the Federal database that matches on this combination, both of the underlying TEAM records will have a matching output of <1> for Sweep 1.
- If there are three TEAM records that have identical versions of the combination for Sweep 3, and there are five records in the Federal database that match on that combination, the three TEAM records will each have a matching output of <2> for Sweep 3.

## Exhibit B

## DMDC Step 1.3 Data Diagnostics

Data Element	Data Availability	"In Texas" Population	"Outside of Texas" Population
Last Name	available	746,010	8,285,517
	missing	0	1
First Name	available	746,004	8,285,426
	missing	6	92
Middle Name	available	669,476	7,656,604
	missing	76,534	628,914
SSN	available	746,010	8,285,517
	missing	0	1
Date of Birth	available	746,010	8,285,518
	missing	0	0
Gender	available	746,003	8,285,379
	missing	7	139
Address Text	available	746,010	8,108,935
	missing	0	176,583
Address State	available	746,010	8,026,896
	missing	0	258,622
Address Zip	available	745,993	8,072,943
	missing	17	212,575

---

Data Element	Data Quality*	"In Texas" Population	"Outside of Texas" Population
SSN	good	744,032	8,211,554
	bad	1,978	73,964
Date of Birth	good	745,936	8,284,603
	bad	74	915

**NOTES:**

An SSN was considered "bad" if it met any of the following criteria:

- \*Equalled 000000000, 111111111, 222222222, 333333333,...,999999999
- \*Equalled sequential values 123456789 or 234567890
- \*The first 3 numbers were 000 or 666
- \*The first number was 9
- \*The middle node was 00 (ie xxx-00-xxxx)
- \*The last node was 0000 (ie xxx-xx-0000)

A Date of Birth was considered "bad" if it met either of the following criteria:

- \*Birth year was 1914 or prior (aka were 100 years of age or older)
- \*Representative of repeating digits (ie 01/01/01, 02/02/02,...,12/12/12)